

Information about deposition

The Repository “CLARIN Centre INL” gives access to language resources and tools from the INL and other CLARIN Members.

Mission

The CLARIN Centre INL stores the resources and tools according to the INL mission:

- (i) To promote the Dutch language, in particular its lexicology and lexicography, through the ages.
- (ii) To conduct scientific research with an emphasis on lexicology, lexicography and related components of computational linguistics.
- (iii) To produce, link and make available source material for Dutch in the form of historical and contemporary corpora, dictionaries, computational lexicons and language technology tools.
- (iv) To manage, maintain and make available electronically via a central portal Dutch source material provided by third parties.
- (v) To provide education in lexicology and (computational) lexicography.

Guidelines for deposition

Data presented for deposition need to be supplied with all information that is essential for sustainable data management and future use.

Moreover, the data should be provided in standard formats endorsed by CLARIN EU (<http://clarin.eu/node/2320>).

For archiving purposes, a minimum set of metadata in valid CMDI format either needs to be provided by the data producer or is extracted by INL from data and documentation.

Data producers are encouraged to supply additional documentation of the data or links to publications (using persistent identifiers) about the data. The publications are stored in the repository (given that permission is granted) while the documentation is archived in the OAI/PMH accessible data repository.

Responsibilities

The data producer will always remain the proprietor. CLARIN Centre INL receives a copy of which it must take good care, according to the terms of the license contract and the terms and conditions for use.

The CLARIN Centre INL also makes copies, for example for the benefit of backup and looks after them well.

In case of an emergency we are able to build up an entire new database composed of all files we backed up and stored safely at another location. Quarterly and yearly tapes are stored at <http://backupned.nl>.

Preservation

To ensure the integrity of the data sets, for every deposited file a checksum (md5 type) is made which allows us to check for defects of the data over the years.

Once deposited, files in data sets are never changed and only minor changes to the metadata are allowed. For example: correction of spelling, minor changes in documentation, additional documentation added. Changes to the data themselves will be issued as a new version of the dataset, which will obtain a new persistent identifier. These changes are only made in close collaboration with the producer of the dataset.

Authenticity

Data producers hand over the materials to us. We do not change the data, except by adding metadata if required.

If applicable, we create collection-level objects which provide a context for the embedded data sets.

The repository maintains links to other relevant materials (e.g. articles, theses, documentation, related data) and to software and tools that have been used in production of the data, if applicable.

The identity of a depositor is ensured by the required login using CLARIN SpF for identification.